Modely pro nestacionární časové řady

1 Nestacionární procesy

1.1 Modely ARIMA

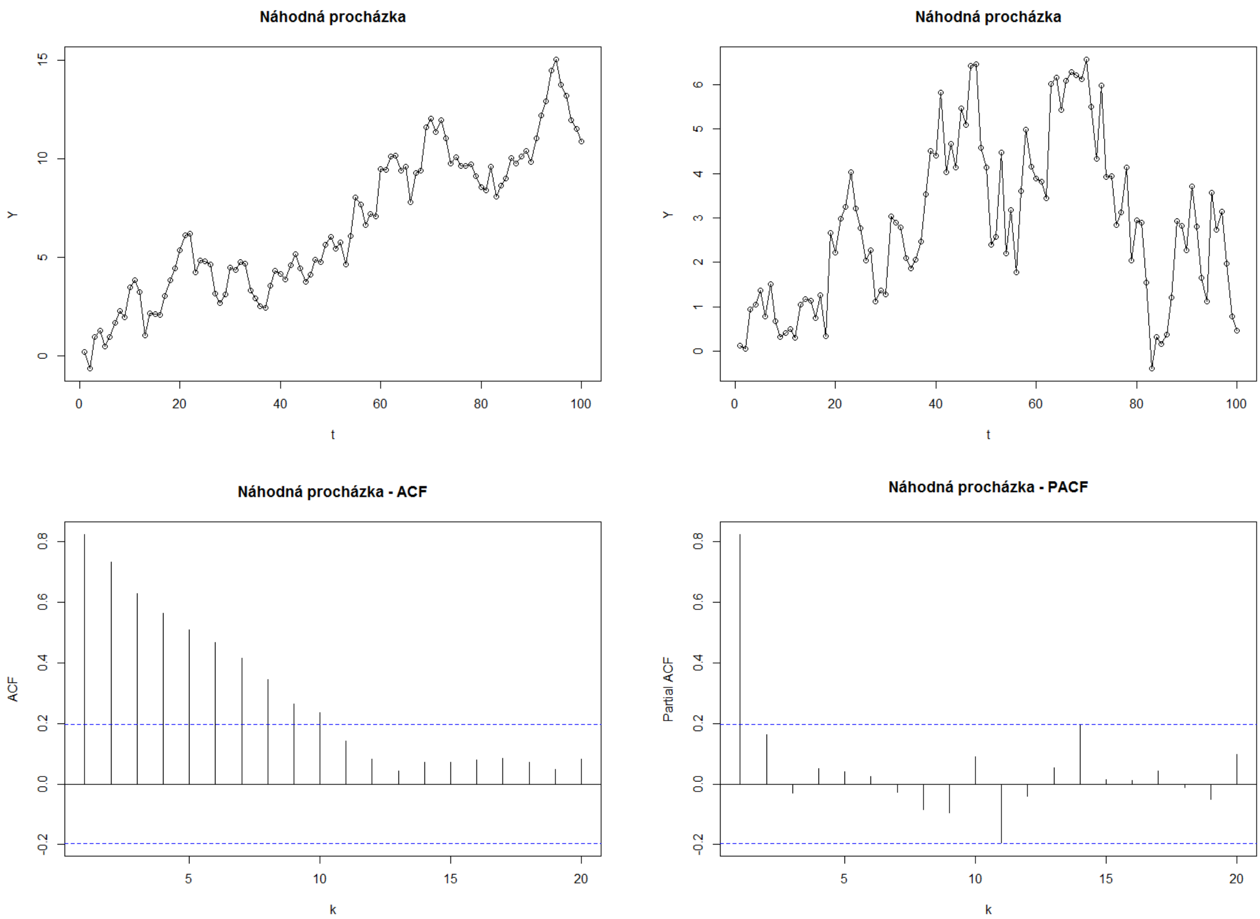
Proces

$$Y_t = Y_{t-1} + \epsilon_t$$

je označuje jako **proces náhodné procházky**. Pomocí operátoru zpětného posunutí lze proces vyjádřit jako

$$(1 - L)Y_t = \epsilon_t.$$

ACF tohoto procesu klesá pomalu, PACF hodnotu $\phi_{11} = 1$, ostatní hodnoty jsou nulové.



Diferenci $\Delta Y_t = Y_t - Y_{t-1}$ lze pomocí operátoru zpětného posunutí zapsat jako

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - LY_t = (1 - L)Y_t.$$

Pro diferenci 2. řádu $\Delta^2 Y_t = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$ lze pomocí operátoru zpětného posunutí zapsat jako

$$\Delta^2 Y_t = (1 - L)^2 Y_t.$$

Operační program Vzdělávání pro konkurenceschopnost
Název projektu: Inovace magisterského studijního programu Fakulty ekonomiky a managementu
Registrační číslo projektu: CZ.1.07/2.2.00/28.0326

PROJEKT JE SPOLUFINANCOVÁN EVROPSKÝM SOCIÁLNÍM FONDEM A STÁTNÍM ROZPOČTEM ČESKÉ REPUBLIKY.

Diferencování časové řady v R-ku provedeme funkcí `diff`.

Pro některý procesy platí, že po transformaci pomocí diference řádu d, je lze popsat jako proces ARMA(p, q). Takový model označujeme jako model ARIMA(p, d, q)

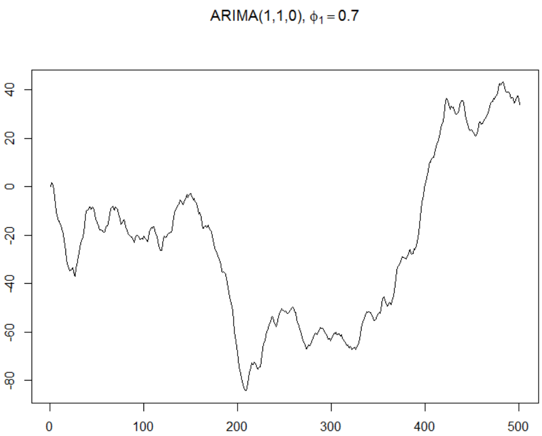
$$(1 - \phi_1 L - \cdots - \phi_p L^p)\Delta^d Y_t = (1 + \theta_1 L + \cdots + \theta_q L^q)\epsilon_t,$$

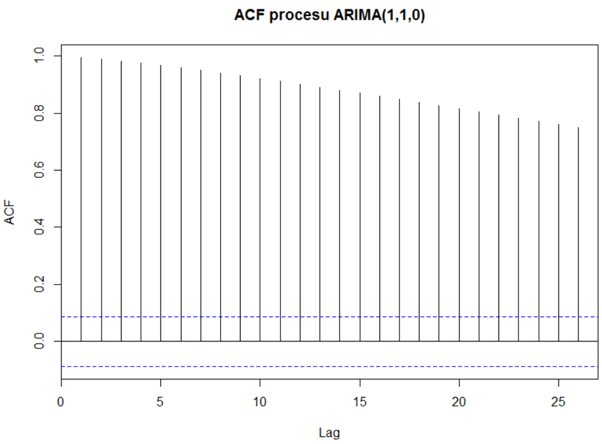
$$(1 - \phi_1 L - \cdots - \phi_p L^p)(1 - L)^d Y_t = (1 + \theta_1 L + \cdots + \theta_q L^q)\epsilon_t.$$

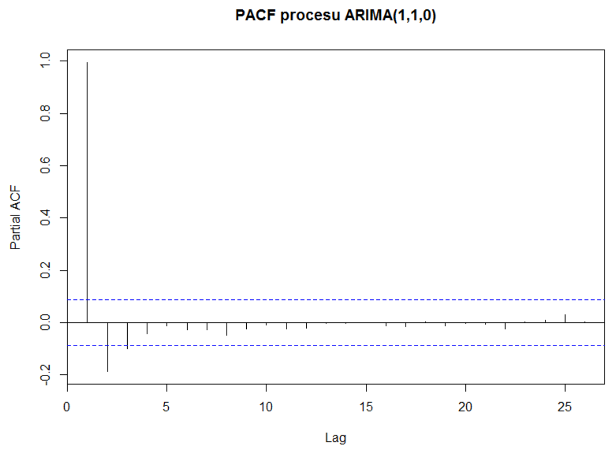
K ověřování nestacionarity procesu slouží tzv. **testy jednotkových kořenů – unit root tests**. Mezi nejznámější patří Dickey-Fullerovy testy (ADF testy).

Odhady parametrů ARIMA modelů získáme v R-ku pomocí funkce `arima`, základní diagnostiku vhodnosti modelu dává funkce `tsdiag`, předpovědi určíme s využitím funkce `predict`.

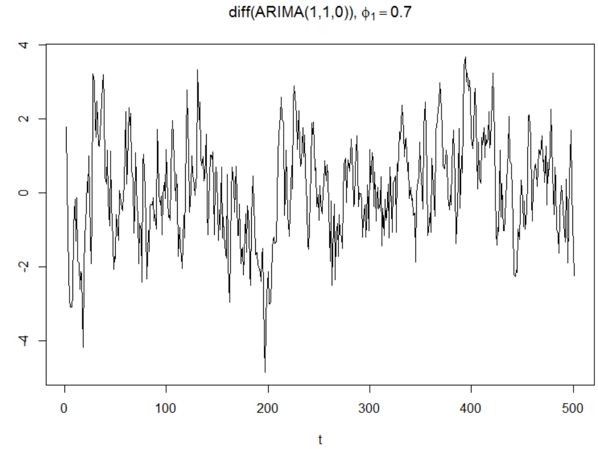
Na následujících grafech je zobrazena realizace procesu ARIMA(1,1,0) s parametrem $\phi_1 = 0{,}7$.. Kromě jí jsou znázorněny i autokorelační a parciální autokorelační funkce.

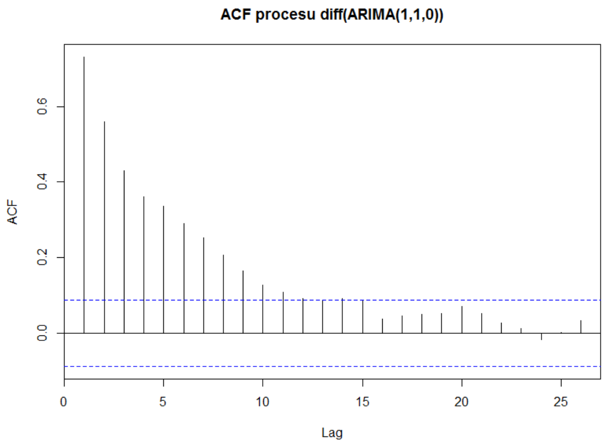


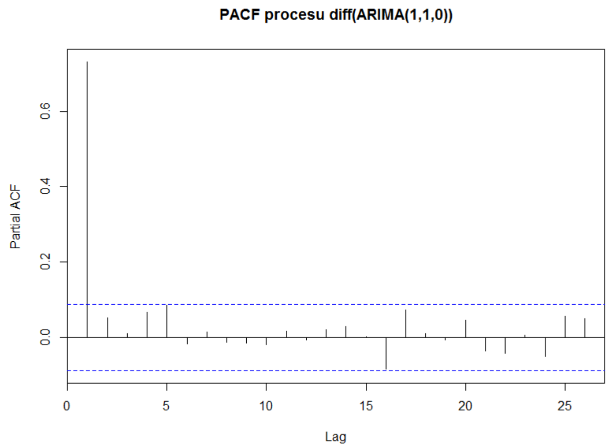




První diference analyzovaného procesu spolu s odpovídající autokorelační a parciální autokorelační funkcí zachyceny na následujících grafech.







1.2 Transformace

Mimo diferencování existují i jiné transformace, pomocí nichž lze dosáhnou stacionarity. Asi nejpoužívanější transformací je logaritmování.

Předpokládejme, že $Y_t > 0$ pro všechna t a že

$$E(Y_t) = \mu_t \quad \text{a} \quad \sqrt{D(Y_t)} = \mu_t \sigma.$$

Předpoklad popisuje situaci, kdy se rozptyl mění v závislosti na střední hodnotě. Potom

$$E(\ln Y_t) \approx \ln \mu_t \quad \text{a} \quad D(\ln Y_t) \approx \sigma^2.$$

Tyto závěry vyplývají z Taylorova rozvoje

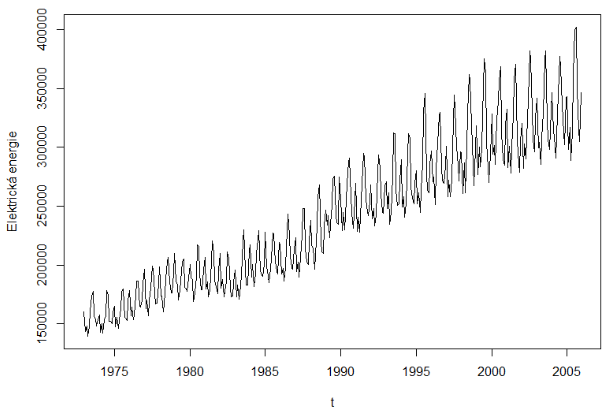
$$\ln Y_t \approx \ln \mu_t + \frac{Y_t - \mu_t}{\mu_t}.$$

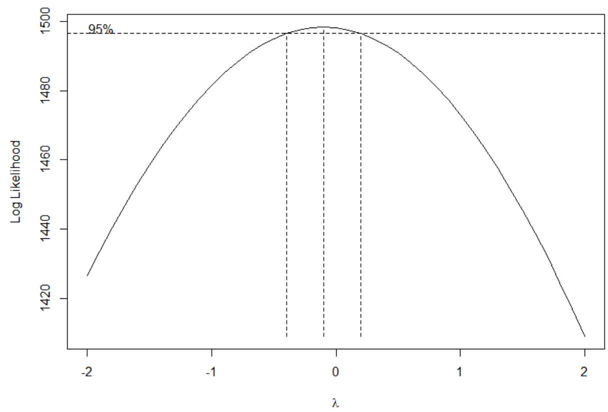
Pro danou hodnotu parametru λ je transformace definována následovně

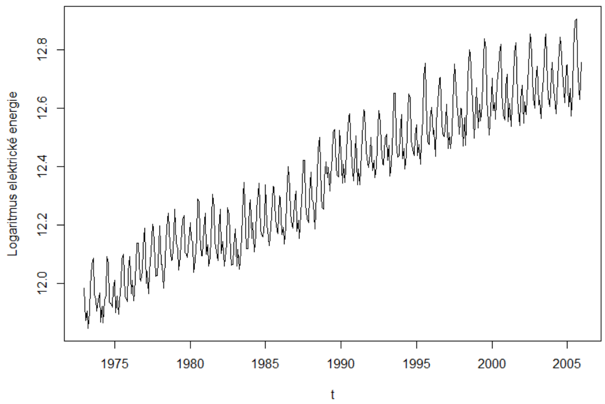
$$g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{pro } \lambda \neq 0, \\ \ln x & \text{pro } \lambda = 0. \end{cases}$$

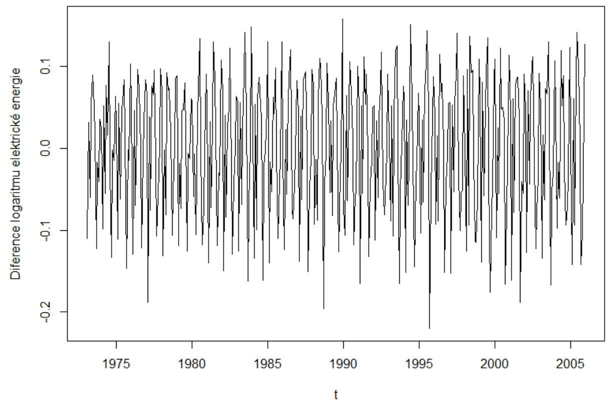
Hodnota parametru λ může být odhadnuta v R-ku pomocí funkce `BoxCox.ar`.

Požití ukážeme na časové řadě popisující množství elektrické energie vyrobené v USA v období 01/1973–12/2005 - měsíční data.



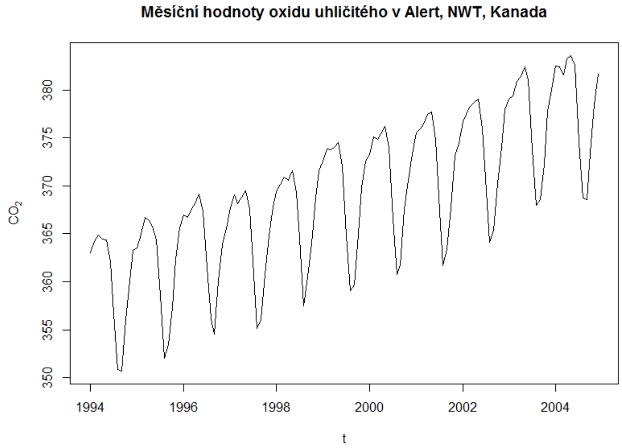






2 Procesy se sezónností

2.1 Modely SARIMA



Uvažujme nejprve stacionární modely. Označme s sezónní periodu (pro měsíční časové řady $s = 12$, pro čtvrtletní $s = 4$). Mějme proces

$$Y_t = \epsilon_t + \Theta\epsilon_{t-12}.$$

Všimněme si, že

$$C(Y_t, Y_{t-1}) = C(\epsilon_t + \Theta\epsilon_{t-12}, \epsilon_{t-1} + \Theta\epsilon_{t-13}) = 0,$$

ale

$$C(Y_t, Y_{t-12}) = C(\epsilon_t + \Theta\epsilon_{t-12}, \epsilon_{t-12} + \Theta\epsilon_{t-24}) = \Theta\sigma_\epsilon^2.$$

Tento proces je stacionární a má nenulové autokorelace pouze pro zpoždění 12. Definujme sezónní MA(Q) proces s periodou s následovně

$$Y_t = \epsilon_t + \Theta_1\epsilon_{t-s} + \Theta_2\epsilon_{t-2s} + \ldots \Theta_Q\epsilon_{t-Qs}.$$

Charakteristický polynom má tvar

$$\Theta(z) = 1 + \Theta_1 z^s + \Theta_2 z^2 s + \cdots + \Theta_Q z^{Qs}.$$

Analogicky definujeme sezónní AR(P) proces s periodou s

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \cdots + \Phi_P Y_{t-2s}$$

s charakteristickým polynomem

$$\Phi(z) = 1 - \Phi_1 z^s - \Phi_2 z^2 s + \ldots - \Phi_P z^{Ps}.$$

Sezónní ARMA model vznikne „spojením“ modelů AR(P) a MA(Q) Sezónní ARMA$(p, q)(P, Q)$ model s periodou s jen model s AR charakteristickým polynomem $\phi(z)\Phi z$ a s MA charakteristickým polynomem

$\theta(z)\Theta(z)$, kde

$$\begin{aligned}
\phi(z) &= 1 - \phi_1 z - \phi_2 z^2 + \ldots - \phi_p z^p,\\
\Phi(z) &= 1 - \Phi_1 z^s - \Phi_2 z^2 s + \ldots - \Phi_P z^{Ps},\\
\theta(z) &= 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q,\\
\Theta(z) &= 1 + \Theta_1 z^s + \Theta_2 z^2 s + \cdots + \Theta_Q z^{Qs}.
\end{aligned}$$

U ARIMA procesů se stacionarity dosáhlo pomocí diferencování ($\Delta Y_t = Y_t - Y_{t-1}$).

U nestacionárních sezónních procesů definujeme **sezónní diferenci**

$$\Delta_s Y_t = Y_t - Y_{t-s}.$$

Lze definovat obecný nestacionární proces SARIMA$(p, d, q)(P, D, Q)$, kde d značí D řád sezónní diference.

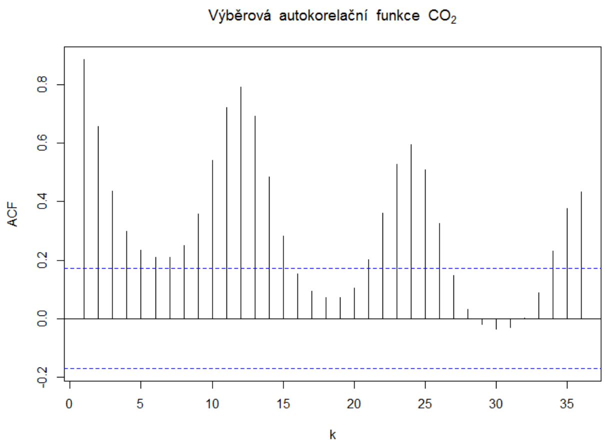
$$\phi(L)\Phi(L^s)\Delta^d\Delta_s^D = \theta(L)\Theta(L^s)\epsilon_t$$

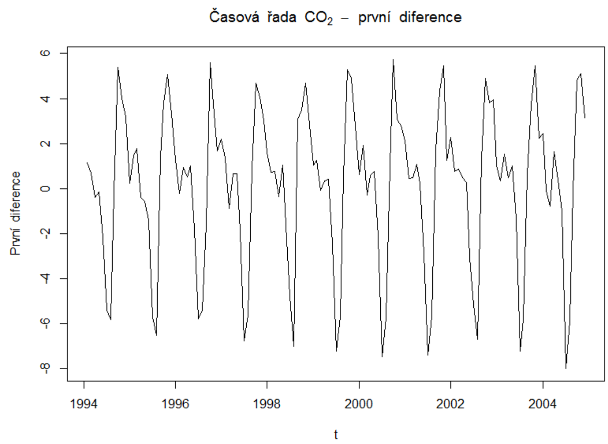
Např. SARIMA$(0, 1, 1)(0, 1, 1)_{12}$ má tvar

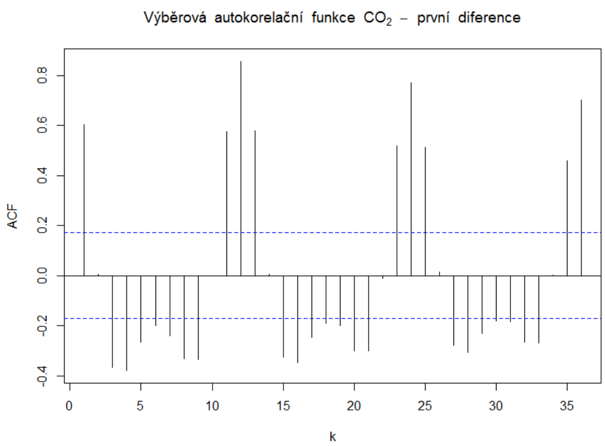
$$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L)\epsilon_t,$$

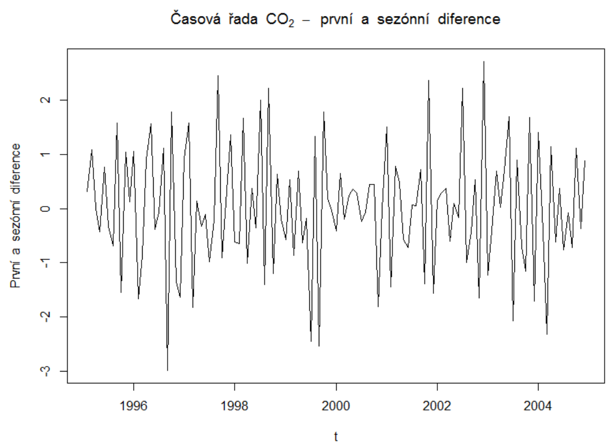
nebo ekvivalentně

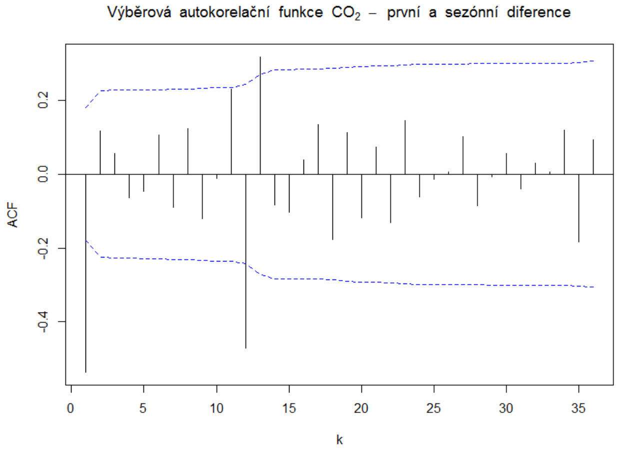
$$Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} = \epsilon_t + \theta_1\epsilon_{t-1} + \Theta_1\epsilon_{t-12} + \theta_1\Theta_1\epsilon_{t-13}.$$



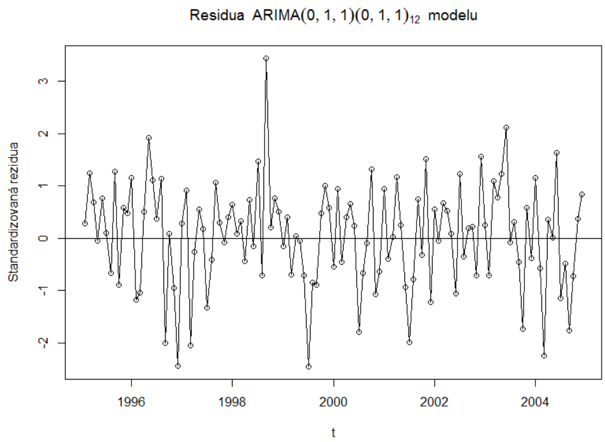


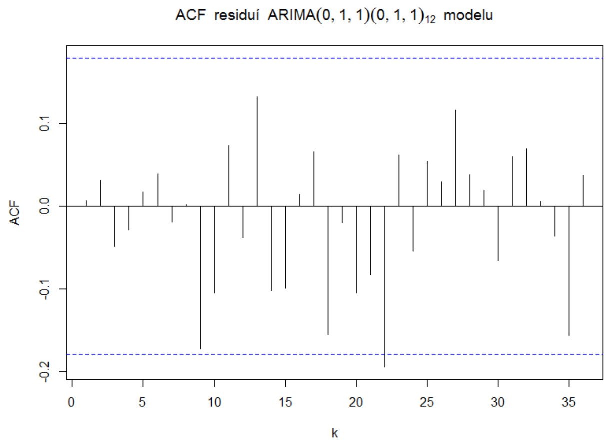


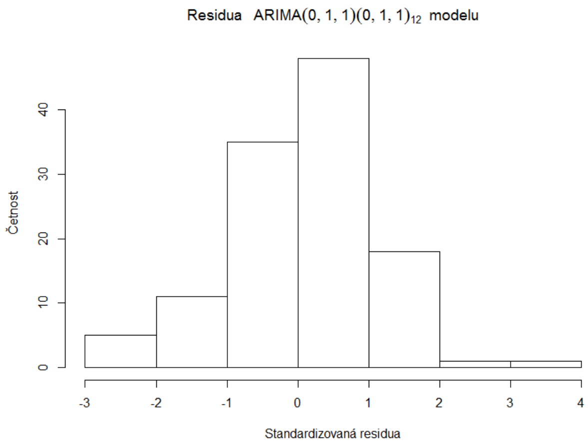


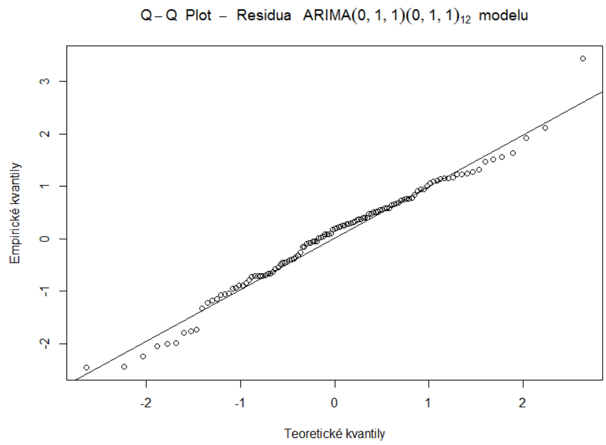


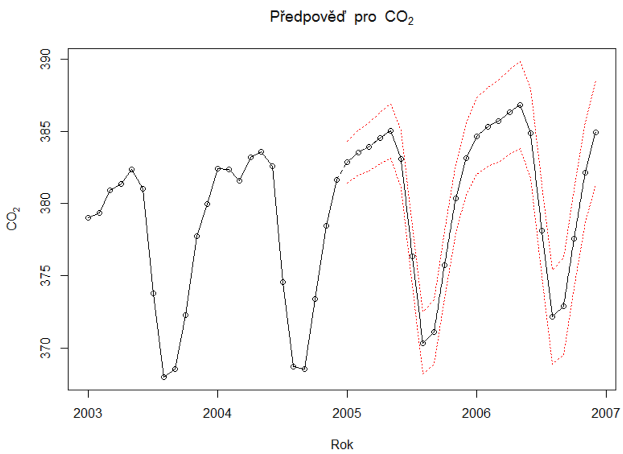

```
Call:
arima(x = co2, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:
          ma1     sma1
      -0.5792  -0.8206
s.e.   0.0791   0.1137

sigma^2 estimated as 0.5446:  log likelihood = -139.54,  aic = 283.08
```











Příklady k procvičení

1. V datovém souboru proces1.txt lze najít časovou řadu délky $n = 200$. Najděte vhodný ARIMA model. Proveďte diagnostiku zvoleného modelu (vlastnosti reziduí) a spočtete predikce 10 budoucích hodnot. [Datový soubor: proces1.txt]
2. V datovém souboru proces2.txt lze najít časovou řadu délky $n = 200$. Najděte vhodný ARIMA model. Proveďte diagnostiku zvoleného modelu (vlastnosti reziduí) a spočtete predikce 10 budoucích hodnot. [Datový soubor: proces2.txt]
3. V datovém souboru proces3.txt lze najít časovou řadu délky $n = 200$. Najděte vhodný SARIMA model. Proveďte diagnostiku zvoleného modelu (vlastnosti reziduí) a spočtete predikce 10 budoucích hodnot. [Datový soubor: proces3.txt]
4. Popište črvrtletní časovou řady indexu spotřebitelských cen od roku 2000 do 2012 (CPI_CR_ctvrtletni.txt) pomocí vhodného pomocí vhodného ARIMA či SARIMA modelu. Proveďte diagnostiku zvoleného modelu (vlastnosti reziduí) a spočtete predikce 10 budoucích hodnot. [Datový soubor: CPI_CR_ctvrtletni.txt]
5. Datový soubor airpass z balíčku „TSA“ obsahuje měsíční údaje o počtu pasažérů na mezinárodních linkách letech 1960 až 1971. Popište tuto časovou řadu vhodným SARIMA modelem. Určete predikce predikce 10 budoucích hodnot. [Příkaz pro R: data(airpass,package=”airpass”)]